Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The DNA-binding domain of BenM reveals the structural basis for the recognition of a T-N_{11}-A sequence motif by LysR-type transcriptional regulators

Amer M. Alanazi,[a]‡ Ellen L. Neidle[b] and Cory Momany[a]*

[a]Department of Pharmaceutical and Biomedical Sciences, University of Georgia, Athens, GA 30602, USA, and [b]Department of Microbiology, University of Georgia, Athens, GA 30602, USA

‡ Current address: Department of Pharmaceutical Chemistry, King Saud University, PO Box 2457, Riyadh 11451, Saudi Arabia.

Correspondence e-mail: cmomany@uga.edu

Received 30 January 2013
Accepted 24 June 2013

PDB References: BenM-DBD/LH, 3m1e; complex with p_{catB} site 1, 4ihs; complex with p_{benM} site 1, 4iht

LysR-type transcriptional regulators (LTTRs) play critical roles in metabolism and constitute the largest family of bacterial regulators. To understand protein–DNA interactions, atomic structures of the DNA-binding domain and linker-helix regions of a prototypical LTTR, BenM, were determined by X-ray crystallography. BenM structures with and without bound DNA reveal a set of highly conserved amino acids that interact directly with DNA bases. At the N-terminal end of the recognition helix (α3) of a winged-helix–turn–helix DNA-binding motif, several residues create hydrophobic pockets (Pro30, Pro31 and Ser33). These pockets interact with the methyl groups of two thymines in the DNA-recognition motif and its complementary strand, T-N_{11}-A. This motif usually includes some dyad symmetry, as exemplified by a sequence that binds two subunits of a BenM tetramer (ATAC-N_7-GTAT). Gln29 forms hydrogen bonds to adenine in the first position of the recognition half-site (ATAC). Another hydrophobic pocket defined by Ala28, Pro30 and Pro31 interacts with the methyl group of thymine, complementary to the base at the third position of the half-site. Arg34 interacts with the complementary base of the 3′ position. Arg53, in the wing, provides AT-tract recognition in the minor groove. For DNA recognition, LTTRs use highly conserved interactions between amino acids and nucleotide bases as well as numerous less-conserved secondary interactions.

1. Introduction

LysR-type transcriptional regulators (LTTRs) control diverse bacterial functions, including antibiotic resistance, CO_2 fixation, nodulation, virulence, aromatic compound degradation and amino-acid biosynthesis (Henikoff *et al.*, 1988; Schell, 1993). LTTRs function as homo-oligomers, usually tetramers, in which each subunit is composed of approximately 300 amino acids. The N-terminal region forms a DNA-binding domain (DBD; residues ∼1–67) with a winged-helix–turn–helix motif. The DBD is connected by a long linker helix (LH; residues ∼70–90) to the effector-binding domain (EBD; residues ∼91–300). The EBD structure, which resembles a periplasmic binding protein, has a cleft between two subdomains in which small regulatory molecules bind. This domain was characterized from CysB (Tyrrell *et al.*, 1997) and OxyR (Choi *et al.*, 2001), after which numerous EBD structures from different LTTRs were described. The first full-length crystal structure, CbnR, was reported a decade ago and since then a few other structures of full-length LTTRs have been characterized (Monferrer *et al.*, 2010; Muraoka *et al.*,

2003; Ruangprasert *et al.*, 2010; Sainsbury *et al.*, 2009; Taylor *et al.*, 2012; Zhou *et al.*, 2010). However, no structure of an LTTR complexed with cognate DNA has been presented.

A common DNA-sequence motif recognized by LTTRs is T-N_{11}-A, although this motif is not universal (Henikoff *et al.*, 1988; Schell, 1993). The two conserved nucleotides are typically within a small region of dyad symmetry. There is matching dyad symmetry in the protein counterpart, as demonstrated by LTTR crystal structures in which LH regions can connect the classic helix–turn–helix (HTH) DNA-binding motifs of two subunits in an appropriate fashion for DNA recognition. As described here, protein–DNA interactions were investigated with BenM, a well characterized LTTR from the soil bacterium *Acinetobacter baylyi* ADP1 (Collier *et al.*, 1998; Craven *et al.*, 2008).

In conjunction with the paralogous LTTR CatM, BenM controls aromatic compound consumption by regulating numerous genes and operons involved in the β-ketoadipate pathway (Collier *et al.*, 1998; Craven *et al.*, 2008). In a segment of this pathway that initiates the conversion of benzoate to tricarboxylic acid cycle intermediates, BenM serves as the primary transcriptional activator of an operon controlled by the *benA* promoter (p_{benA}). At this locus, CatM also activates transcription, but at levels that are too low to support growth on benzoate as a sole carbon source without BenM. In contrast, CatM activates high-level transcription from the *catB* promoter (p_{catB}) controlling another operon involved in the same pathway. At p_{catB}, BenM activates low-level transcription that is insufficient for rapid growth on benzoate without CatM.

The genes encoding *benM* and *catM* are upstream of, and divergently transcribed from, *benA* and *catB*, respectively (Fig. 1). Similarity in the aligned operator–promoter region reflects that the DNA-recognition portions of BenM and CatM are 82% identical (DBD residues 1–60). In the absence of inducers, a BenM tetramer binds two palindromic sites such that expression from p_{benA} is repressed (site 1 and site 3, consensus A**T**AC-N_7-GT**A**T; Fig. 1; Bundy *et al.*, 2002). The position of BenM is shifted by two inducers, benzoate and its catabolite *cis*,*cis*-muconate (hereafter designated muconate). Benzoate and muconate cause conformational changes that result in a tetramer binding to site 1 and another sequence, site 2, which differs from the consensus recognition site by one nucleotide (Fig. 1). This shift uncovers the −10 region of the promoter and enables BenM to activate transcription. Without inducers, CatM similarly binds to site 1 and site 3 of p_{benA}. Muconate causes a shift to sites 1 and 2 to allow low-level transcriptional activation. However, unlike BenM, CatM does not respond to benzoate (Ezezika *et al.*, 2006; Craven *et al.*, 2008).

The response to benzoate depends on its binding to a hydrophobic pocket of the BenM EBD that is not present in the CatM EBD despite the similarity of their crystal structures (Ezezika *et al.*, 2007; Craven *et al.*, 2009). This hydrophobic pocket is distinct from a cleft between two EBD subdomains that serves as the central binding site for effector molecules. Muconate can bind in this cleft (termed the primary binding site) of both the BenM EBD and the CatM EBD (Ezezika *et al.*, 2007). To evaluate the atomic-level changes responsible for transcriptional activation, crystal structures of full-length BenM and two variants were compared. While a single amino-acid replacement (E226K) produces a regulator that activates transcription without benzoate or muconate, the full-length structure of this variant was nearly identical to that of wild-type BenM (Ruangprasert *et al.*, 2010). Therefore, the structural determinants of the effector and promoter specificity of BenM and CatM remain unclear.

The site 1 sequences of both p_{benA} and p_{catB} have perfect ATAC-N_7-GTAT palindromes that appear to anchor BenM and CatM to the promoters. However, the sequences of the site 2 and site 3 regions deviate from the consensus and differ between the two operator–promoter regions. Here, we describe crystal structures of the portion of BenM involved in DNA recognition (BenM-DBD/LH) alone and in complexes in which two subunits are bound to the site 1 region of p_{benA} or p_{catB}. These protein–DNA structures are the first to reveal the molecular basis of sequence-specific recognition by an LTTR.

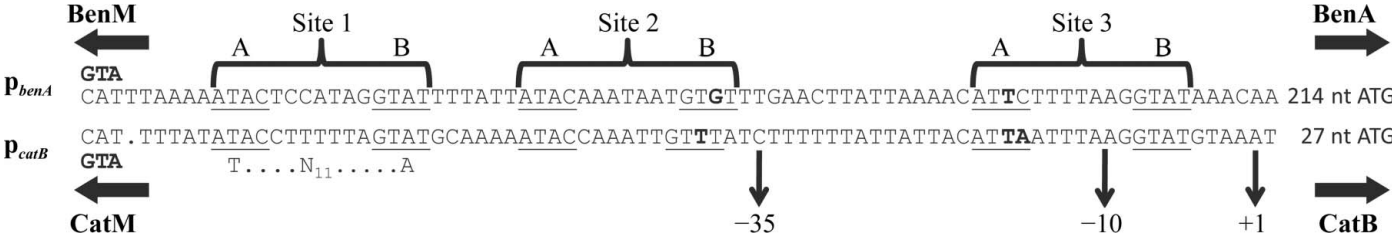

Figure 1
The operator–promoter regions of two operons: *benABCDE* (p_{benA}) and *catBCIJFD* (p_{catB}). This alignment compares the top strand of DNA governing the divergent expression of *benM* and *benA* with that for the divergently oriented *catM* and *catB*. The transcriptional start sites for *benA* and *catB* (+1) are shown, as are the positions of the promoter elements (−10 and −35). BenM is the primary regulator for p_{benA} where, in the absence of effectors (benzoate and muconate), a tetramer represses *benA* by binding sites 1 and 3. Effectors shift the BenM position to sites 1 and 2 to activate high levels of *benA* expression. CatM can similarly regulate p_{benA} but is only capable of low-level transcriptional activation and only responds to muconate. As is common for LTTRs, *benM* and *catM* are both negatively autoregulated. At p_{catB} CatM activates high-level transcription in response to muconate, whereas at this promoter BenM activates only low-level transcription in response to muconate. A similar model for DNA binding is proposed for p_{catB} as for p_{benA} (Craven *et al.*, 2008). The labelled binding sites (1, 2 and 3) share similar dyad symmetry (ATAC-N_7-GTAT) discussed in terms of half-sites (A and B) each able to bind one DBD of a tetramer subunit. The bold text shows deviations from perfect symmetry. The half-sites of sites 2 and 3 are not well conserved, and one half-site (B) of the p_{catB} site 2 is likely to be shifted one nucleotide as indicated by the underlining.

The results should apply broadly to this large and diverse family because the DBD regions of LTTRs are highly conserved.

2. Materials and methods

2.1. Cloning, purification and crystallization of BenM-DBD/LH

An expression vector (pBAC952) was made that encodes the DBD and LH of BenM (residues 1–87) with a C-terminal hexahistidine purification tag. To generate this plasmid, the DNA encoding the EBD was excised from the full-length gene (on pBAC433). This excision was achieved by PCR mutagenesis using 5′-phosphorylated primers followed by self-ligation of the gel-purified PCR product. The forward primer introduced a new glycine residue into the protein sequence before the C-terminal hexahistidine purification tag. The construct was verified by sequencing and was transformed into *Escherichia coli* BL21(DE3) RIL cells (Stratagene Inc.) for protein expression. The transformed cells were grown overnight at 310 K in 100 ml autoinduction medium (Studier, 2005) and were harvested by centrifugation at 7000*g* for 10 min at 277 K; the pellets were then resuspended in 12 ml binding buffer (20 m*M* Tris, 0.5 *M* NaCl, 20% glycerol, 10 m*M* *β*-mercaptoethanol, 5 m*M* imidazole pH 8). The suspended cells were lysed using a pre-chilled French pressure cell at 110 MPa. The cell lysate was then centrifuged at 39 000*g* for 30 min at 277 K and the supernatant was loaded onto a 1 ml HisTrap metal-chelate column (GE Biosciences) charged with Ni^{2+} and equilibrated in the binding buffer. The protein was eluted using a linear gradient to binding buffer containing 500 m*M* imidazole. The purified BenM-DBD/LH protein was dialyzed into 20 m*M* Tris, 0.1 *M* NaCl, 10% glycerol, 10 m*M* *β*-mercaptoethanol, 0.5 m*M* EDTA pH 8.0. It was then concentrated to 7 mg ml^{-1} using Millipore Ultrafree centrifugal concentrators at 277 K for use in crystallization. Unliganded BenM-DBD/LH protein was crystallized *via* the microbatch under paraffin oil method (Chayen, 1997) at 288 K using crystallization screening kits purchased from Hampton Research (Index). Crystals were obtained by mixing 2 µl concentrated protein in the metal-chelate elution buffer with an equal volume of 0.1 *M* HEPES pH 7.5, 10% PEG 6000, 5% MPD. The protein concentration was determined using a Bio-Rad protein assay colorimetric microassay with BSA as the standard.

2.2. DNA-complex preparation and crystallization

Crystallization trials of BenM-DBD/LH with DNA were performed with DNA sequences differing in their overhangs and lengths (Table 1). DNA-fragment design was based on DNase I footprinting studies (Bundy *et al.*, 2002). The results of these studies suggested that 25 bp fragments would interact well with the DBDs. Crystallization trials included the p_{catB} site 1 and 2 sequences. Several permutations of the lengths of the oligonucleotides were explored using unpurified oligonucleotides prepared on the minimum commercially available scale (25 nmol scale; purchased from IDT). Complementary oligonucleotides (0.1 m*M* each) in 10 m*M* Tris, 10 m*M* NaCl, 1 m*M* EDTA pH 7.5 were annealed to make DNA duplexes by heating to 368 K on a heating block followed by slow cooling to room temperature. To prepare protein–DNA complexes, the BenM-DBD/LH protein was incubated with the annealed DNA duplexes separately for 30–45 min on ice at protein dimer:DNA duplex molar ratios of 1:1.2. The complexes were then concentrated to various degrees dependent on their solubility behavior. The complex of BenM-DBD/LH with p_{catB} site 1 DNA was concentrated to 20 mg ml^{-1} (protein concentration), while the complex of BenM-DBD/LH with p_{benA} site 1 DNA was concentrated to 9.6 mg ml^{-1} (protein concentration). Crystallization trials of the DNA–protein complexes utilized reagent-specific crystallization kits (Grid Screen Ammonium Sulfate, Grid Screen PEG 6000, Grid Screen Sodium Malonate and Grid Screen Sodium Chloride) from Hampton Research. A minimalistic screen of conditions favorable for nucleic acid protein complexes was manually explored using the microbatch method.

Table 1
Oligonucleotide sequences used in the crystallization screening trials.

DNA	Sequence†	Base pairs
p_{benA} site 1	5′-TAAAAA**T**ACTCCATAGGT**A**TTTTAT 3′-ATTTTT**A**TGAGGTATCC**A**TAAAATA	25
p_{benA} site 2	5′-TTATTA**T**ACAAATAATGTGTTTGAA 3′-AATAAT**A**TGTTTATTACACAAACTT	25
p_{catB} site 1	5′-TTTATA**T**ACCTTTTTAGTATGCAAA 3′-AAATAT**A**TGGAAAAATCATACGTTT	25
p_{catB} site 2	5′-CAAAAA**T**ACCAAATTGTTTATCTTT 3′-GTTTTT**A**TGGTTTAACAAATAGAAA	25
p_{catB} site 1 short	5′-TATA**T**ACCTTTTTAGT**A**TGCA 3′-ATAT**A**TGGAAAAATC**A**TACGT	21
p_{catB} site 1 short-overhang	5′- ATA**T**ACCTTTTTAGT**A**TGCA 3′-TTAT**A**TGGAAAAATC**A**TACG	20

† Nucleotides corresponding to the T-N_{11}-A recognition motif used by LTTRs are shown in bold. The 5′-ATAC-3′ half-site sequence recognized by BenM and its complement are underlined.

The crystallization conditions used in the trials were centered around neutral pH based on crystallization results for other protein–DNA complexes (Dock-Bregeon *et al.*, 1999; Joachimiak & Sigler, 1991). In general, five conditions (at most) from each kit with different concentrations around neutral pH were sampled and 50 conditions were prepared overall. If a mixture precipitated immediately, a lower concentration of precipitate was used. Crystals were only obtained with p_{benA} site 1 and p_{catB} site 1 DNA. Crystals of BenM-DBD/LH with p_{catB} site 1 DNA were grown from 2.4 *M* sodium malonate pH 6.0 using AI's oil (Hampton Research). Optimal crystals were also obtained from precipitants consisting of 2.4 *M* ammonium sulfate, 0.1 *M* MES pH 6.0 and of 1.6 *M* ammonium sulfate, 0.1 *M* citric acid pH 5.0. BenM-DBD/LH complexed with p_{benA} site 1 DNA crystallized from

Table 2
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	BenM-DBD/LH, PDB entry 3m1e	BenM-DBD/LH–p_{benA} site 1, PDB entry 4iht	BenM-DBD/LH–p_{catB} site 1, PDB entry 4ihs
Beamline†	19-BM	19-BM	22-ID
Wavelength (Å)	1.0080	0.9793	0.9999
Resolution range (Å)	50–1.8 (1.86–1.80)	50–2.9 (2.95–2.90)	200–3.1 (3.21–3.10)
Space group	$C222_1$	$P2_12_12$	$P4_22_12$
Unit-cell parameters			
a (Å)	33.360	58.956	156.493
b (Å)	84.146	300.295	156.493
c (Å)	68.734	46.003	141.525
$\alpha = \beta = \gamma$ (°)	90	90	90
Total reflections	45207	65096	250855
Unique reflections	9190	15997	30319
Multiplicity	4.9 (3.0)	3.7 (3.3)	8.2 (3.0)
Completeness (%)	98.91 (92.78)	92.80 (96.77)	92.05 (71.20)
Mean $I/\sigma(I)$	23.75 (2.92)	26.13 (6.35)	25.65 (4.36)
Wilson B factor (Å^2)	23.73	81.27	84.80
R_{merge}‡	0.045 (0.360)	0.065 (0.503)	0.114 (0.399)
R factor	0.1850 (0.2906)	0.1845 (0.2399)	0.1782 (0.2408)
R_{free}§	0.2340 (0.3379)	0.2452 (0.3237)	0.2029 (0.2569)
No. of atoms	818	9001	9126
Macromolecules	720	4910	4960
Ligands	1	0	2 malonate, 1 Na^+
Water	97	3	20
No. of protein residues	87	455	464
R.m.s.d., bonds (Å)	0.006	0.008	0.007
R.m.s.d., angles (°)	0.92	1.24	1.10
Ramachandran favored (%)	100	93	96
Ramachandran outliers (%)	0	0	1.1
B factor (Å^2)			
Average	32.0	95.6	109.3
Protein	32.4	90.0	100.7
DNA	—	104.4	122.0
Solvent	29.7	78.9	80.3

† At the Advanced Photon Source, Argonne, Illinois, USA. Beamline 19-ID was operated by the Structural Biology Consortium Collaboratory Access Team (SBC-CAT) and beamline 22-ID by the Southeast Regional Collaboratory Access

one condition: 25 m*M* bis-tris pH 6.5, 25% PEG 3000, 50 m*M* ammonium acetate.

2.3. X-ray data collection, structure determination and refinement

Diffraction data were collected and processed on SBC-CAT beamline ID-19 and SER-CAT beamline ID-22 at the Advanced Photon Source, Argonne, Illinois, USA as described previously (Ezezika *et al.*, 2007; Ruangprasert *et al.*, 2010). Data from the SER-CAT beamline were collected remotely with software developed by SER-CAT. The results from the best of several data sets that were collected for each crystal form are reported (Table 2). The structural determination of the unbound BenM-DBD/LH crystals used the *CCP*4 package (Winn *et al.*, 2011) and molecular replacement using the DBD/LH of CbnR (residues 1–100 of chain *A*; PDB entry 1ixc; Muraoka *et al.*, 2003) as the search model in the program *AMoRe* (Navaza, 1994). CbnR and BenM are 50% identical in sequence in the DBD region. After rigid-body refinement of helices within the HTH domain and the LH, an atomic model was built into the calculated electron density that revealed appropriate sequence differences between BenM and CbnR. Refinement of the structures was performed using *REFMAC*5 v.5.5.0072 (Murshudov *et al.*, 2011) with TLS domains interspersed with cyclic model building and water identification using *Coot* (Emsley & Cowtan, 2004). TLS domains were identified by the *TLSMD* server (http://skuld.bmsc.washington.edu/~tlsmd/; Painter & Merritt, 2006). H atoms were included in the refinement in the last few cycles (R_{free} decreased 2%). The final model included amino acids 1–87 encoded by the *benM* gene. No electron density was visible for the polyhistidine tail.

The crystals of the protein–DNA complexes did not diffract to as high a resolution as the unbound crystal (Table 2). The crystals of both DNA complexes showed marked defects in the diffraction images, with the BenM-DBD/LH–p_{catB} site 1 diffraction spots clearly split. The processing software was successful in integrating the intensities despite the diffraction defects, and analysis of the data during data reduction did not reveal any underlying twinning problems. A BenM-DBD/LH dimer was generated by applying twofold crystallographic symmetry to the unliganded BenM-DBD/LH structure and was used as the search model for the BenM-DBD/LH–p_{catB} site 1 complex using the various space-group permutations possible with the 4/*mmm* Laue symmetry. The best molecular-replacement solution positioned two dimeric units in the $P4_22_12$ space-group cell. Electron density for the DNA was readily visible in difference Fourier maps phased from the protein alone. A variety of DNA structures were extracted from the Protein Data Bank, and the model best fitted by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) was changed to the correct sequence. Amino acids up to residue 86 had electron-density features in all structures, and all side chains were included in refinement. Only in the *C* subunit of the protein–p_{benA} structure was residue 87 not modeled. Histidine residues and a glycine linker associated with the C-terminal polyhistidine purification tail had variable degrees of electron density and were omitted from individual subunits when their calculated electron densities were not visible in $2F_o - F_c$ and $F_o - F_c$ maps. All residues of the DNA had clear electron density. *REFMAC* refinement parameters for the nucleic acids were modified to match the published parameters (Gelbin *et al.*, 1996; Clowney *et al.*, 1996). A round of refinement with

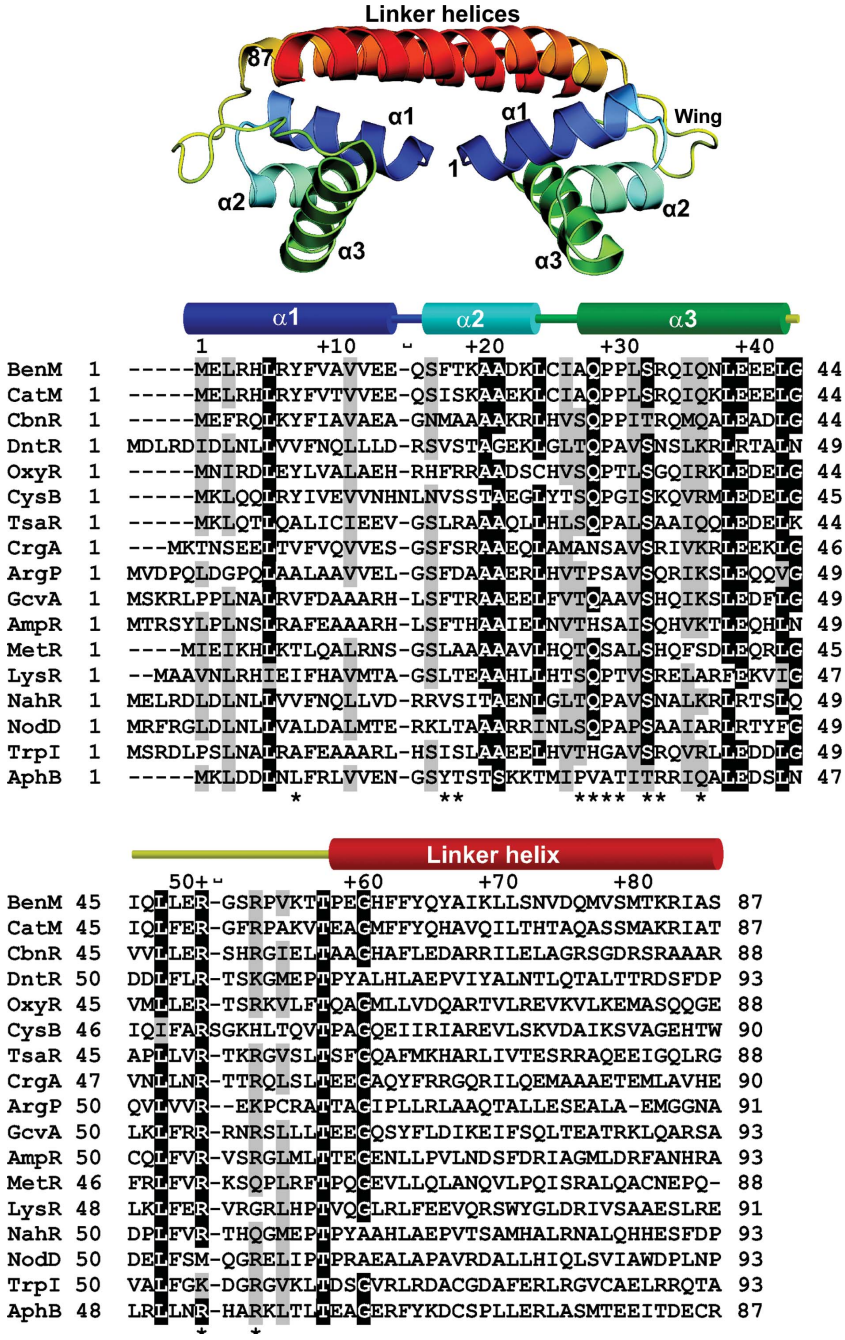

Figure 2
Structure of BenM-DBD/LH and amino-acid sequence alignment of selected LTTRs in the DBD/LH region. Shown at the top is a ribbon representation of the BenM-DBD/LH dimer coloured blue to red from the amino-terminus to the carboxy-terminus. In the sequence alignment below, highly conserved residues among the LTTRs are highlighted in black and gray, while the three helices of the DBD are represented by cylinders above the sequence. Numbers above the alignment correspond to the amino-acid positions within BenM. Brackets indicate gaps in the BenM sequence. Critical residues discussed in the text are denoted with an asterisk below the sequences. The alignment was created with the program *BioEdit* using aligned output from *ClustalX*. The protein designations, organisms and UniProt sequence identifiers are BenM, *Acinetobacter baylyi* ADP1, O68014; CatM, *A. baylyi* ADP1, PO774; CbnR, *Ralstonia eutropha*, Q9WXC7; DntR, *Burkholderia* sp. DNT, Q7WT50; OxyR, *E. coli*, P0ACQ4; CysB, *Salmonella typhimurium*, P06614; TsaR, *Comamonas testosteroni* T-2, P94678; CrgA, *Neisseria meningitides*, Q9JPU9; ArgP, *Mycobacterium tuberculosis*, P67665; GcvA, *E. coli*, P0A9F6; AmpR, *Enterobacter cloacae*, P05051; MetR, *S. typhimurium*, P0A2Q4; LysR, *E. coli*, P03030; NahR, *Pseudomonas putida*, P10183; NodD, *Rhizobium meliloti*, P03031; TrpI, *P. aeruginosa*, P11720; AphB, *Vibrio cholera*, Q9XC54. Proteins were selected based on their presence in the Protein Data Bank, their biological importance and the degree to which they have been studied.

PHENIX (Adams *et al.*, 2010; Afonine *et al.*, 2012) was used to introduce H atoms, define TLS groups and orient Gln, Asn and His side chains appropriately before the final rounds of *REFMAC* (Murshudov *et al.*, 2011) refinement. The BenM-DBD/LH–p_{benA} site 1 structure was similarly solved by molecular replacement as described above. An electron-density feature at the interface of the two dimeric units that was not easily interpreted was modeled as two malonate molecules with a single sodium ion acting as a bridge between the carboxylates of the malonate. Figures were prepared using *PyMOL* (DeLano, 2002). The web-based versions of the programs *Curves+* (Blanchet *et al.*, 2011; Lavery *et al.*, 2009) and *3DNA* (Zheng *et al.*, 2009) were used to analyze the DNA structures. The *DaliLite* v.3 web server was used to identify protein structural neighbors (Holm & Rosenström, 2010).

3. Results

3.1. Crystallization, structural determination and structure of BenM-DBD/LH

BenM-DBD/LH was expressed and purified with a C-terminal hexahistidine purification tag. Crystallization screens identified several conditions that produced crystals of this protein. The highest resolution data set came from crystals obtained using the precipitant 0.1 *M* HEPES pH 7.5, 10% PEG 6000, 5% MPD. Molecular replacement using the DBD and LH of CbnR as the model provided adequate phases for modeling and refinement of the complete structure of BenM-DBD at 1.8 Å resolution (Table 2).

The single molecule of BenM-DBD/LH in the asymmetric unit forms the biologically relevant two-subunit binding region of the tetramer by application of crystallographic twofold symmetry. As expected from the high sequence similarity among the DBDs of LTTRs, the BenM-DBD structure shares structural features, such as the classic HTH motif, with the DBDs of CbnR (Muraoka *et al.*, 2003) and those in the full-length structures of CrgA (Sainsbury *et al.*, 2009), TsaR (Monferrer *et al.*, 2010), ArgP (Zhou *et al.*, 2010) and AphB (Taylor *et al.*, 2012). The BenM-DBD

consists of helix α1 (residues 1–15), a turn region (residues 16–17), helix α2 (residues 18–25), a second turn region (residues 26–28), helix α3 (residues 29–43), a 'wing' (residues 44–58) and the LH (residues 59–83), as indicated in Fig. 2. The position of the wing is different from the classic winged HTH in which the wings normally directly interrupt, precede or follow the HTH motif (Huffman & Brennan, 2002; Parkinson *et al.*, 1996; Ramakrishnan, 1997; Clark *et al.*, 1993; Clubb *et al.*, 1994; Harrison *et al.*, 1994). As first observed in the structure of CbnR, the wing lies between the DNA-recognition helix and the LH of the protein (Muraoka *et al.*, 2003). As expected with 1.8 Å resolution data, there was good overall electron density of the unliganded structure. However, the BenM-DBD wing had weaker electron density with respect to the rest of the structure and less clearly defined secondary structure in that region. The structure of this region differs from the CbnR-DBD wing, which has a clear β-hairpin conformation.

3.2. Crystallization and structural determinations of BenM-DBD/LH complexes with DNA

Like many other DNA-binding proteins, BenM-DBD binds to a unique inverted DNA-repeat sequence. We chose to crystallize oligonucleotides with authentic promoter sequences that include site 1 or site 2 of p_{benA} and p_{catB}. While each sequence contains ATAC, they are not symmetric overall. In the site 1 region, the palindromic sequences of p_{benA} and p_{catB} are the same (ATAC-N_7-GTAT). However, BenM normally activates high-level transcription at p_{benA} and only low-level transcription from p_{catB}, so structural differences in these protein–DNA complexes could be significant. Only site 1 duplexes crystallized with BenM-DBD/LH. The lack of perfect dyad symmetry in the binding sequences of site 2 DNA could account for weaker interactions with BenM-DBD/LH and the failure to crystallize. *In vivo*, interactions at site 2 or site 3 are likely to be stabilized by interactions of the full-length oligomer with site 1 DNA. The p_{benA} site 1 DNA complex crystallized in space group $P2_12_12$ with two BenM-DBD/LH dimers and DNA duplexes in the asymmetric unit.

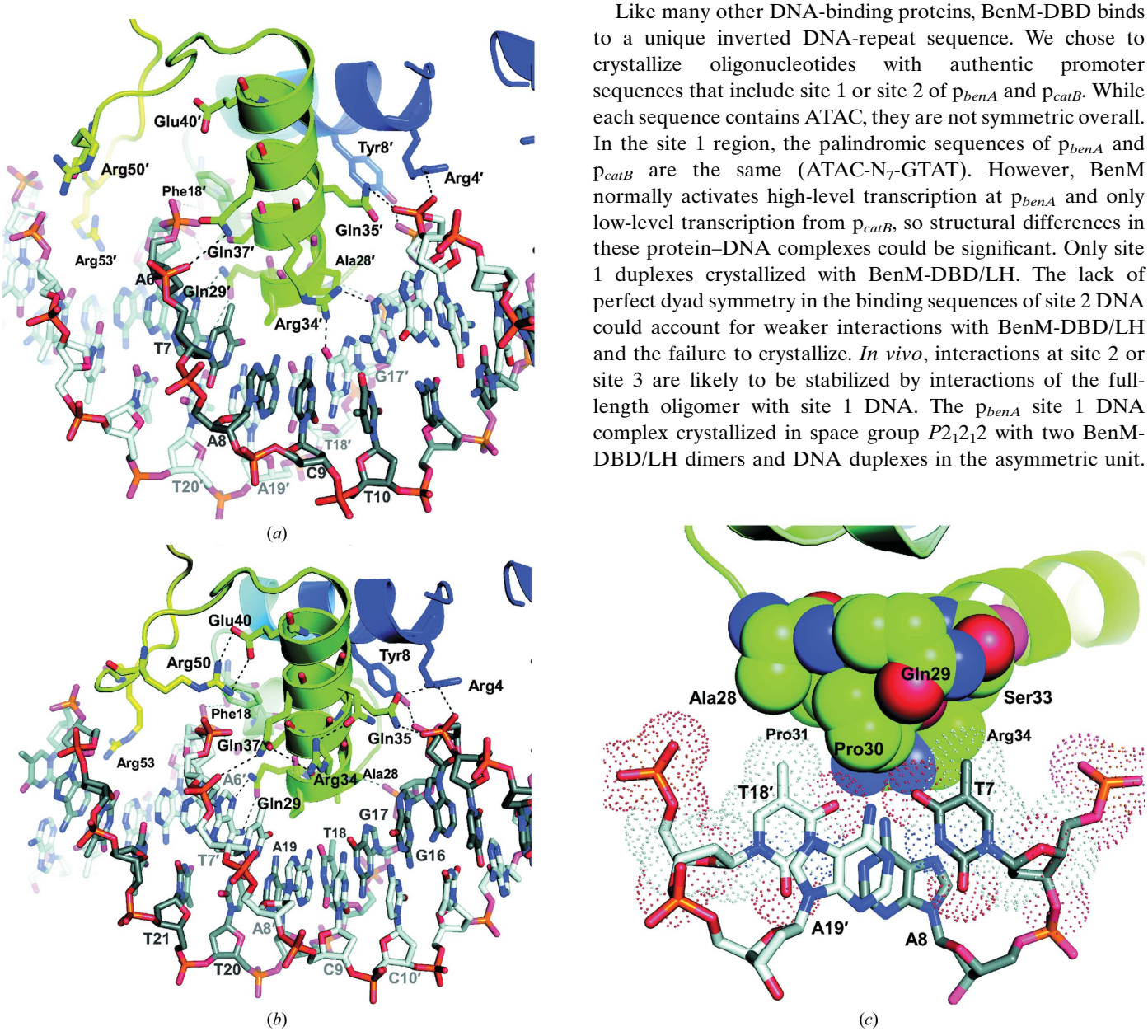

Figure 3
BenM-DBD interactions with p_{benA} site 1 DNA. (*a*) Residues at the N-terminal end of the recognition helix α3 and the wing of a BenM-DBD subunit (chain *B*) interact with DNA in the p_{benA} complex. The color scheme is as in Fig. 2. Selected hydrogen-bond interactions between the protein and DNA are represented as dotted lines. The C atoms of chain *F* are colored dark gray, while the chain *E* C atoms are white. Arg34 interacts with bases in the DNA. (*b*) A similar view as in (*a*) of the *A*-chain protein residues. In this chain, Arg34 projects away from the DNA and hydrogen bonds to Gln37 and Gln35, which directly interact with phosphates from the DNA. (*c*) The methyl groups of two thymine bases pack tightly against residues from the recognition helix. Amino acids from BenM-DBD are shown as CPK spheres with van der Waals radii. Dots represent the van der Waals surfaces of the bases.

The p_{catB} site 1 DNA complex crystallized with two dimeric units of BenM-DBD/LH and DNA duplexes in the asymmetric unit of crystals belonging to space group $P4_22_12$. Interestingly, the crystal packing of the BenM-DBD/LH subunits created a tight interaction with protein face-to-face contacts reminiscent of the hydrophobic surfaces of natural protein–protein interactions. This feature of BenM-DBD/LH may explain why there are no other DBDs of LTTRs in the Protein Data Bank as it could contribute to aggregation of the polypeptide, a problem that we encountered with CatM-DBD/LH (unpublished results). In the p_{benA} site 1 structure the DNA packs end-on to form continuous superhelices in the crystal along the long crystallographic *c* axis. In the p_{catB} site 1 structure, end-on DNA packing is present, but continuous helices are not created along the *c* axis. Instead, continuous DNA helices run perpendicular to the *c* axis along the *a* and *b* cell axial directions.

3.3. General structural features of BenM-DBD/LH with site 1 DNA from p_{benA} or p_{catB}

In these structures, the two recognition helices of one BenM-DBD/LH dimeric unit occupy two consecutive turns of the major groove per DNA duplex. In many other transcriptional regulators, the two recognition helices of one dimer occupy a more continuous region of the major groove (Xu *et al.*, 2001). The solvent-accessible surface areas between the BenM dimeric unit and the double-stranded DNA helices were 1719 and 1800 Å^2 on average for p_{benA} site 1 and p_{catB} site 1, respectively. Remarkably, only a few water molecules were found in these protein–DNA complex structures and no waters were involved in bridges between the protein and the DNA. As is commonly seen in protein–DNA complexes, the bound DNA is significantly bent. The p_{benA} site 1 DNA molecules have cumulative bend angles of 47° (*EF* subunits) and 45° (*GH* subunits) calculated with the program *Curves+* (Blanchet *et al.*, 2011; Lavery *et al.*, 2009), while the p_{catB} site 1 DNA molecules are bent at angles of 47° (*EF* subunits) and 46° (*GH* subunits) (see Supplementary Material[1]). Between the ATAC recognition half-sites (nucleotides 6–20 of each chain), the cumulative bend angles are 38° for the p_{benA} *EF* subunits and 42° for the *GH* subunits; those for the p_{catB} subunits are 40° and 37°, respectively. The largest bend angles between base-pair steps occur consistently in all of the structures between the recognition half-sites, with a maximum bend of 3.9° between the $A_{15}T_{16}$ and $G_{16}C_{17}$ steps of the p_{benA} site 1 *GH* subunits. The range of larger bend angles (around 3° for each step) is expanded into the second half-site in the GH duplex of the p_{benA} DNA, where Arg34 is pulled away from the DNA. When withdrawn, Arg34 orients two residues, Gln37 and Asn38, to interact directly with the DNA, which introduces a slightly sharper local bend relative to the other DNA duplex structures. The overall bend difference is cancelled out by lower contributions to the total bend angle at the ends of the *GH* helix. Most other deviations in the nucleic acid conformations are minor and are within the range of normal B-DNA conformations. The largest base-pair step roll angles (see Supplementary Material) occur between the half-sites (maximum difference of −10.9° in the *benA EF* subunits) as well as the first base-pair step outside the ATAC recognition half-site sequences in the A/T-rich regions of *benA* site 1 (−11.6° at the $T_{20}T_{21}$ step in the *benA GH* subunits) and the first half-site of *catB* site 1 (−9.6° for the T_5A_6 step).

The electron density of the wing is better defined in the protein–DNA complexes than it is for the BenM-DBD/LH protein alone. Nevertheless, the definition of the side chains is poor in this region and some modeling ambiguity remains. Because this protein–DNA interaction is located at the ends of the oligonucleotides used for crystallization, there may be end effects that could be responsible for weakening the interaction of the wing with the DNA. Arg50 and Glu40, which are shown in Fig. 3, help position the wing *via* a salt bridge, an interaction that is not observed in the unliganded structure. An additional hydrogen bond to the carbonyl O atom of Leu48 locks the guanidino group of Arg50 in place. In the unbound structure, Arg50 is exposed to the solvent and interacts with Glu40 through neighboring water molecules. Most of the reported full-length structures, including CbnR and TsaR, also lack this critical salt bridge. Furthermore, on superimposition of these other structures onto the DNA-bound BenM-DBD/LH-structures, their wings would not contact the DNA. In most of the BenM-DBD/LH subunits the wing forms an oval shape rather than an extended β-hairpin, and it makes several contacts with the DNA phosphate backbone. In some subunits, but not all, Arg53 projects into the minor groove of the A-rich regions flanking the recognition sequences. The position of Arg53 is marked in Fig. 3. Recognition of the minor grove by arginine residues has been observed in multiple DNA-binding proteins associated with poly-A tracts (Rohs *et al.*, 2009).

For the p_{benA} DNA, the width of the minor groove is minimal at the fifth (2.5 Å in the *EF* subunits and 2.2 Å in the *GH* subunits) and 21st (2.7 Å in the *EF* and *GH* subunits) nucleotides in the DNA (one nucleotide upstream and downstream from the ATAC sequence), where Arg53 intercalates into the DNA minor groove. In other subunits (*G* and *H*; BenM-DBD–p_{catB} site 1), Arg53 is modeled interacting with the phosphate of residue A25 in the DNA. There are also two conformations of Arg4, a residue that is labeled in Fig. 3. Arg4 either interacts with the phosphate group of adenine 15 or is projected away from the phosphate group.

The DNA-binding surface of BenM has many hydrogen-bond donors and positively charged amino acids, making it ideal for association with the negatively charged phosphodiester backbone of DNA (Figs. 4 and 5). Positive charge along the DNA backbone comes from two arginine side chains (Arg4 and Arg34; shown in Fig. 3), two lysine side chains (Lys20 and Lys24 in helix α2) and the previously mentioned wing residues Arg53 and Arg50. The dipole moment of helix α2 is also well aligned to provide an interaction with the DNA backbone around the phosphate of A6. Additional hydrogen

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5228). Services for accessing this material are described at the back of the journal.

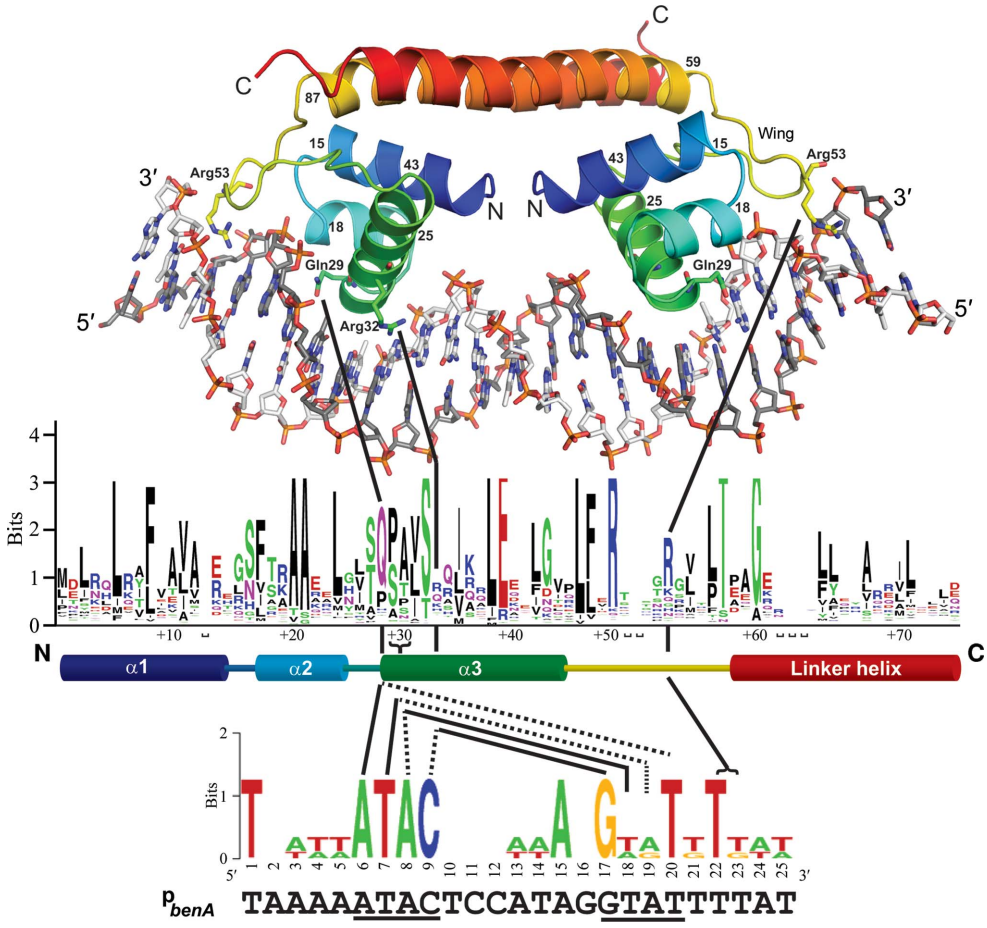

Figure 4
Structure of BenM-DBD/LH bound to p_{benA} site 1 DNA and sequences that are conserved in other LTTRs. The ribbon representation of the p_{benA} BenM DBD/LH dimer is colored as in Fig. 2. Each BenM-DBD subunit consists of three helices, α1, α2 and α3, followed by a wing that precedes the LH. Below the structure image is a multiple sequence alignment of LTTRs visualized with the program *WebLogo* (http://weblogo.berkeley.edu/). The amino-acid sequences used to produce the alignment representation were all proteins that were annotated as having the LysR-type HTH-domain profile (PS50931) in the Prosite database with a 50% identity cutoff to reduce redundant sequences and to better represent the diversity of the profile. Amino-acid numbering follows the scheme used in Fig. 2. At the bottom is a *WebLogo* representation of the three DNA-binding sites in p_{benA} and the p_{benA} sequence (site 1) in the crystal structure. The reverse complement sequence for site 3 was used in the *WebLogo* alignment so that the conserved ATAC is on the left-hand side of the alignment, while the sequence diversities of sites 2 and 3 (see Fig. 1) are represented on the right-hand side of the sequence. Solid lines extending from the amino-acid sequence represent base-specific interactions with the strand of DNA with the sequence shown, while dashed lines represent amino-acid interactions with the complementary strand of DNA. To simplify the representation only a subset of interactions is shown.

bonds bridging the protein to the DNA backbone come from the amino-acid side-chain atoms (–OH and –NH_2) of Tyr8, Ser17, Thr19, Gln29, Ser33, Gln35 and Gln37 and the main-chain amide N atoms of Phe18 and Ala28. Some of these residues are labeled in Fig. 3. The side chain of Phe18 also makes van der Waals contacts with the phosphate backbone and deoxyribose.

In general, DNA recognition involves direct readout, a term that describes specific contact between protein side chains and nucleotide bases. Another process, indirect readout, can additionally contribute to promoter specificity. Indirect readout refers to the effects of local DNA differences that cause sequence-dependent deformations of the deoxyribose-phosphate backbone that control protein recognition (Brennan & Matthews, 1989; Otwinowski *et al.*, 1988). Based on our structures of the BenM–DNA complexes, every base associated with the conserved recognition sequence ATAC is in direct contact with amino acids from the protein. Thus, while indirect readout may provide additional specificity, direct readout appears to be the principal mechanism of DNA recognition for BenM, CatM and other LTTRs using conserved sequences (Fig. 4).

3.4. The molecular basis of the T-N_{11}-A DNA motif recognized by many LTTRs

The structures of BenM-DBD/LH with site 1 DNA of p_{benA} and p_{catB} clearly explain the molecular basis of the T-N_{11}-A DNA-binding motif that is prevalent among the known DNA-binding sites of LTTRs. Most of the sequence-specific interactions between BenM-DBD and site 1 DNA are common to both crystal structures because both site 1 DNA sequences contained two ATAC half-sites. A total of 27 amino acids of a BenM-DBD dimer make 36 contacts with the DNA duplex (Fig. 5).

The major DNA sequence-specific interactions seen in the structures are between the N-terminal residues of the recognition helix α3 and the ATAC half-site duplex. The most critical protein residues in base-specific recognition are Ala28, Gln29, Pro30, Pro31 and Arg34, all of which are clustered on the recognition helix α3 (many are labeled in Fig. 3). All of these residues project into the major grove of the DNA. Working from the 5′ end of the half-site (*i.e.* from the 5′ end of p_{benA}), the first adenine base of the **A**TAC sequence interacts with Gln29 through hydrogen bonds from the $O^{\varepsilon1}$ and $N^{\varepsilon2}$ atoms of Gln29 to the N6 and N7 atoms of the adenine base. The side chain of glutamine is ideal for recognition of adenine owing to the combination of a hydrogen-bond donor and an acceptor that is accessible in the major groove at adenine bases. This glutamine is remarkably conserved in LTTR sequences (Figs. 4 and 5). The next nucleotide encountered, T7 (in A**T**AC),

interacts *via* van der Waals interaction of its methyl group with a hydrophobic pocket defined by the side chains of Pro30 (C^{α} and C^{β} atoms) and Ser33 (C^{β} atom) and, in some subunits, with Arg34 (Fig. 3). After this point, protein interactions occur with the complementary DNA chain, with the thymine methyl group of the complement of A8 (AT**A**C) interacting with a second hydrophobic pocket defined by the side chains of Ala28, Pro30 and Pro31, with the C^{γ} and C^{δ} atoms providing the predominant surface interactions from the prolines. The van der Waals interactions are tight around the methyl groups

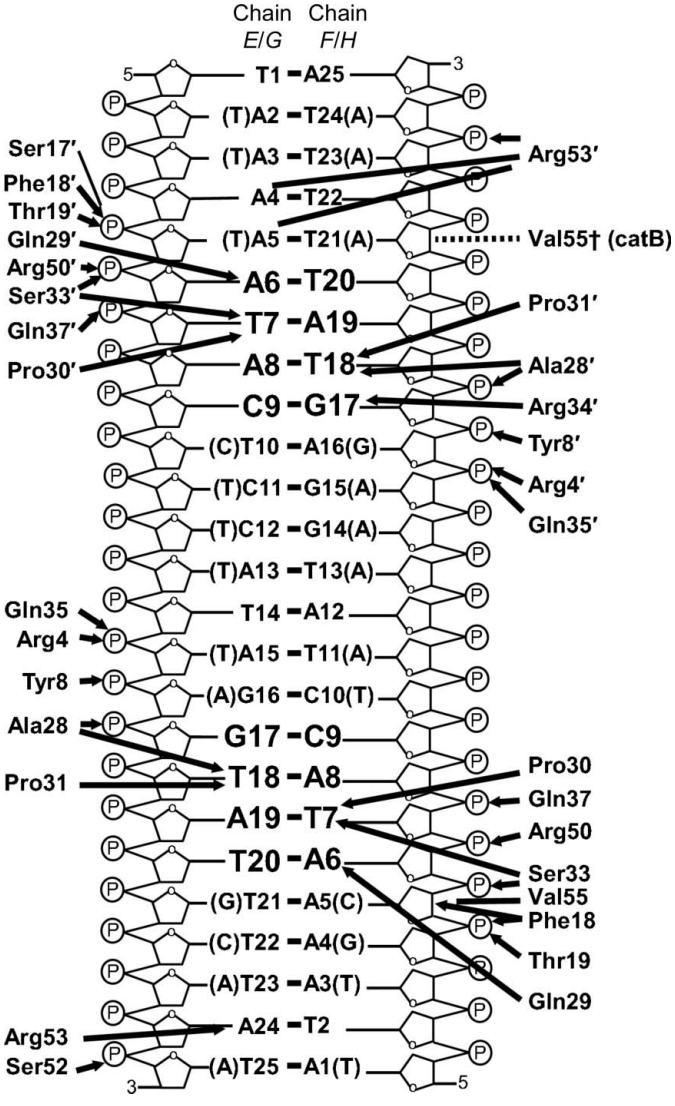

Figure 5
BenM-DBD protein–nucleotide interaction map. The sequence shown running vertically in the center represents the DNA sequence of the p_{benA} site 1 with the *E*/*G* chains on the left and the complementary chains of *F*/*H* on the right. Nucleotides in p_{catB} that differ from p_{benA} are shown in parentheses. The lines from amino-acid residues to the nucleotide label denote sequence specific contacts with the base of the nucleotide, lines to the carbohydrate backbone schematic denote interactions with atoms in the deoxyribose ring and arrows to the circled P denote interactions with the phosphate backbone O atoms. Amino acids from the *A*/*C* protein subunits lack primes, while amino acids from the *B* and *D* subunits are denoted with primes.

of the two thymines, creating excellent complementary surfaces, as shown in Fig. 3(*c*).

In most of the subunits the guanidino side chain of Arg34 contacts the major groove in the general region of G17′, the complement of C9 (ATA**C**). The surface of the major groove in this region tends to have multiple hydrogen-bond acceptors available for interaction with the arginine side chain. For instance, in several subunits of the p_{catB} site 1 complex the NH atoms of Arg34 interact directly with the carbonyl O atom (O6) of guanine G17′. In other subunits, the Arg34 side chain contacts the N7 atom of A16′ (p_{benA}) and the N7 atom of G17′ (p_{catB}). Surprisingly, in the p_{benA} complex (subunits *A* and *C*), Arg34 instead forms a salt bridge to the neighboring Glu41 (up the α3 helix away from the DNA) and hydrogen bonds to Asn38 and Gln37, both of which contact the phosphate backbone in some subunits. A protein–protein crystal contact is nearby that may be responsible for changing the local interaction schemes. However, if the dynamic nature of Arg34 is biologically relevant, the interaction network might represent a conformational switch that could be associated with promoter recognition or RNA polymerase activation.

Residues that may be involved in indirect readout of sequence-positioned phosphates include the previously discussed amino acids Ala28, Ser33, Gln35, Gln37, Asn38 and Arg50. The amide N atom of Ala28 is hydrogen bonded to the phosphate group of G17′ (ATA**C**). Gln37 $N^{\varepsilon 2}$ is hydrogen bonded to the phosphate group of T7 (A**T**AC). The hydroxyl O atom of Ser33 and the ω-NH of Arg50 are hydrogen bonded to the phosphate of A6 (**A**TAC). Overall, these residues may clamp the protein against the DNA for a sensitive readout of the DNA, in particular to enhance the relatively weak van der Waals attraction associated with the two thymine methyl groups within the recognition sequence.

The side-chain methyl group of Thr19 contacts the thymine methyl group of T5 (*E*/*G* chains of p_{catB}). Interestingly, thymidine deoxynucleotides are only positioned comparably to T5 in the site 2 and site 3 anchoring half-sites of p_{benA} and only in site 1 of p_{catB} (Fig. 1). Since threonine is not highly conserved among LTTR amino-acid sequences, the threonine–thymidine interaction may be important in promoter specificity. For example, subtle differences are expected to distinguish the functions of BenM and CatM at similar operator–promoter regions. Indeed, there are only five amino-acid differences in the HTH motif of these two regulators, and this threonine is replaced by a serine in CatM (Fig. 2). A second difference between CatM and BenM is adjacent: Phe18 in BenM *versus* Ile18 in CatM.

3.5. Sequence conservation in DNA recognition by LTTRs

Many of the amino acids in the BenM-DBD are well conserved among different LTTRs (Figs. 2 and 4). Interestingly, the strongest sequence conservation appears to be in residues that play structural roles. As an example, Ala22, which is almost absolutely conserved, functions to maintain the orientation of the two helices with respect to one another and it is distant from the DNA. For the remainder of this

discussion, we focus on DNA-binding residues that are common across LTTRs.

The T-N_{11}-A motif recognition is dependent on Pro30, Ser33 and Arg34 in BenM. Pro30 is remarkably well conserved among LTTRs, with only similarly short side-chain replacements such as serine, alanine and glycine being tolerated at this position. Changes in the residues at positions structurally equivalent to Pro30 of BenM-DBD cause drastic reductions in DNA binding in another LTTR, NahR (Schell *et al.*, 1990). Ser33 is substantially conserved in LTTRs and plays a dual role: methyl-group recognition of the thymidine *via* its C^{β} atom and hydrogen bonding *via* O^{γ} to the phosphate O atom of A6. CbnR contains a threonine at this position, which should be compatible with the dual role. In fact, threonine might enhance binding by filling a slight void that is visible in our structures with its extra methyl group. Gln29 is frequently found in LTTRs of diverse bacteria. Together, the Gln29 interaction and thymidine-binding pockets suggest a common level of promoter specificity for most LTTR family members that recognize the DNA sequence ATA. The neighboring Ala28 is not well conserved among LTTRs, but replacements by serine and threonine residues are structurally feasible and the addition of a hydroxyl to Ala28 would provide enhanced interaction with the phosphate backbone. Ser32, which is remarkably well conserved, is clearly important. Replacement (S33N) in OxyR of *E. coli* abolishes DNA binding (Kullik *et al.*, 1995). The corresponding replacements S34R in CysB (Colyer & Kredich, 1994) and S38P in GcvA (Jourdan & Stauffer, 1998) similarly cause loss of DNA binding.

According to our DNA-complex structures, the wing interacts with the minor groove. The minor-groove regions of both p_{benA} site 1 and p_{catB} site 1 have upstream A tracts that could create narrow minor grooves optimal for electrostatic interaction of the arginine guanidinum group. Arg53 plays this role in BenM, with assistance from Arg50 and Glu40, which help to position the wing (Fig. 3). Supporting the important role of Arg50 in DNA recognition, when the corresponding amino acid is changed to a tryptophan residue in a variant of *E. coli* OxyR, DNA binding is abolished (Kullik *et al.*, 1995). We anticipate that many LTTR recognition sites will include poly-A runs near the recognition sequences because of the high conservation of the wing features and the retention of Arg53 in many LTTRs (although not all) or an arginine near it in the amino-acid sequence (Figs. 2 and 4). Despite the co-occurrence of Arg50 and Glu40 in many LTTR sequences, the amino-acid sequences are otherwise variable among LTTRs in this region. Intriguingly, the greatest divergence between BenM and its paralog CatM occurs in the wing and in residues that cluster around it and near the N-terminal end of the LH. Specific residues in the wing may play a subtle role in recognition specificity.

In one LTTR from *Mycobacterium tuberculosis*, ArgP, strong electron-density differences exist between the open and closed conformations of the DBD subunits around the residues equivalent to Arg50 and Glu40 in BenM (Zhou *et al.*, 2010). Open and closed LTTR structures may reflect conformational changes that occur in the functional oligomers upon binding effector molecules. In ArgP, Arg55 pairs with Glu45 in the closed conformation, but it is exposed towards the solvent in the open conformation. The two conformations of Arg55 in the DBDs were proposed to be owing to the different chemical environments (Zhou *et al.*, 2010). Similar distinctly different subunit conformations have also been reported in the full-length structure of CbnR (Muraoka *et al.*, 2003). Moreover, in the structure of TsaR the electron density of the wing basic residues Lys53 and Arg54 in the extended protomer are defined, whereas in the compact subunit they are disordered (Monferrer *et al.*, 2010). Two subunit conformations are a common feature of full-length LTTR structures, with the sole exception being CrgA, which has only one conformation (Sainsbury *et al.*, 2009). The structural differences between the unbound and bound structures of BenM-DBD/LH suggest that the wing assists in appropriate positioning of the DNA-recognition helix. The correctly positioned helix allows minor-groove interactions between the DNA phosphate backbone and the wing residues. When coupled to conformational differences in subunits, sequence-dependent specificity can be achieved that in turn may be dependent on ligand binding, as proposed for CbnR (Muraoka *et al.*, 2003) and ArgP (Zhou *et al.*, 2010).

3.6. Relationship to other DNA-binding proteins

Specific hydrogen bonds between nucleotides and DNA-binding proteins have been characterized in many studies (Brennan & Matthews, 1989; Dock-Bregeon *et al.*, 1999; Harrison & Aggarwal, 1990). Therefore, some of the hydrogen-bonding patterns observed in BenM are expected, *e.g.* those between adenine and Gln29 and between guanine and Arg34. Interestingly, these two residues in the DNA-recognition helix contact the major groove very similarly to residues from proteins that, while having overall different folds, position their recognition helix in a comparable fashion to BenM. Examples that illustrate protein–DNA contact similar to that of BenM include SlyA, P22, Cro repressor, MosR, TetR and RovA (Fig. 6 and Supplementary Fig. S1).

In the structure of an SlyA–DNA complex (PDB entry 3q5f), Gln60, Pro61 and Arg65 occur in the same relative positions of the recognition helix as the DNA-binding residues Gln29, Pro30 and Arg34 of BenM (Fig. 6*a*; Dolan *et al.*, 2011). Arg65 in SlyA contacts the guanine base of G14 (chain *C*), but this base is offset one nucleotide from the contact in BenM. Gln60 of SlyA is turned away relative to Gln29 in BenM and creates a hydrophobic pocket for a thymine methyl group. Ser33 of BenM may contribute to the difference in the orientation of Gln29 owing to a hydrogen bond that forms between its hydroxyl group and the side chain of Gln29. The same Gln60, Pro61 and Arg65 combination is observed in the protein–DNA structure of RovA, a homolog of SlyA. In RovA (Quade *et al.*, 2012), two residues (Arg78 and Asp68) function similarly to Arg50 and Glu40 in BenM by stabilizing the wing *via* salt bridging, a type of stabilization that is seen in many other protein–DNA complex structures. In other structures, the combination of a glutamine, a proline and an

arginine does not exactly match that of BenM, yet other comparably positioned residues make similar contacts with DNA. For example, in MosR (Brugarolas *et al.*, 2012), one arginine (Arg75) is positioned like Arg34 of BenM, whereas a second arginine (Arg70) has a side chain that interacts with guanine O6 and N7 atoms and acts similarly to a glutamine in BenM (Gln29).

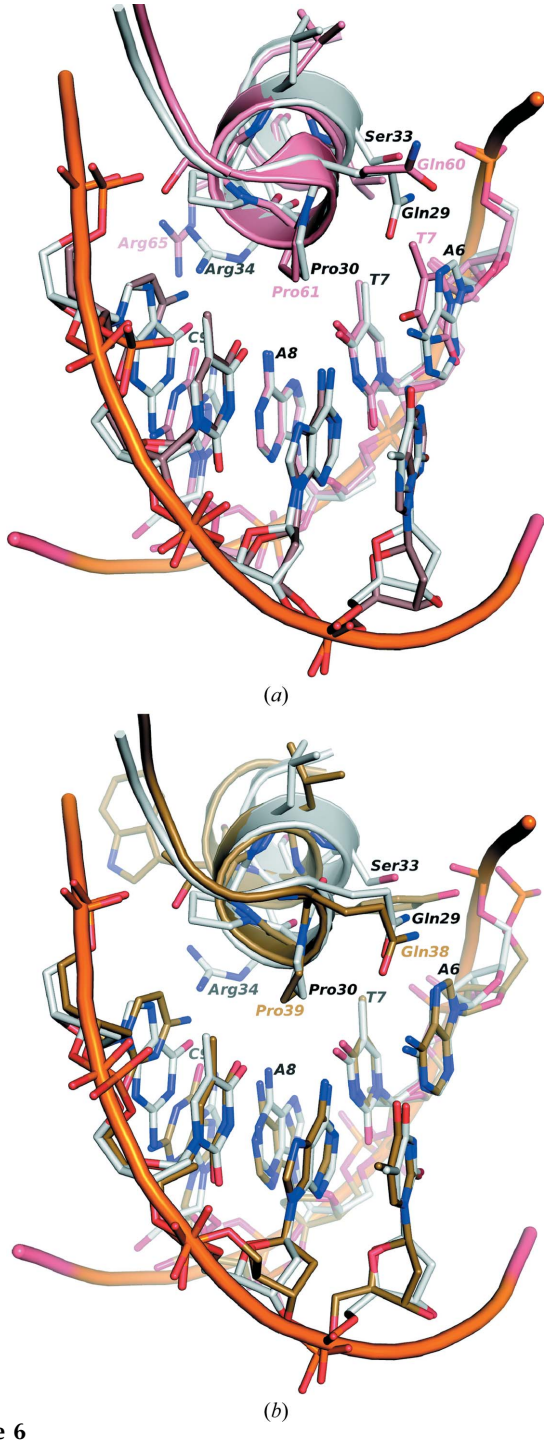

Figure 6
Similarity between BenM-DBD and other DNA-binding proteins in the recognition-helix interactions with DNA. Protein–DNA complexes of (*a*) SlyA (PDB entry 3q5f) and (*b*) TetR (PDB entry 1qpi) were aligned with the amino-acid backbone atoms of BenM-DBD (residues 29–34 of chain *B*) and residues 6–9 of chain *E* and residues 17–20 of chain *F* using the align function of *PyMOL* v.1.3. C atoms from the p_{benA}–BenM-DBD/LH complex are gray, while C atoms of the aligned proteins are colored salmon (SlyA) and gold (TetR). Residues in the text are identified by nearby labels colored using the same color scheme as the C atoms. Additional similarly aligned protein–DNA complexes are shown in Supplementary Fig. S1.

In addition to hydrogen bonds, van der Waals interactions can be used in direct-readout base contacts. However, such contacts tend to be under-reported. For example, proline is not listed in an online nucleic acid–protein interaction database (http://aant.icmb.utexas.edu/aint_top.html) of base–side-chain contacts (Hoffman *et al.*, 2004). Nevertheless, van der Waals interactions were reported in some of the first protein–DNA structures, such as that of the Cro repressor (Anderson *et al.*, 1981). In the P22 repressor, a valine at the end of the recognition helix defines a binding surface that provides recognition of a TTAA DNA duplex through van der Waals interactions (Watkins *et al.*, 2008). When the HTH motif of P22 in the DNA-bound structure is aligned with that of BenM-DBD, this valine matches the side-chain position of Pro30 of BenM. The previously discussed proteins RovA, MosR and SlyA, as well as tetracycline repressor, TetR (Fig. 6 and Supplementary Fig. S1), all use a proline side chain to define a hydrophobic pocket for recognition of a TA DNA duplex in a similar manner to BenM.

TetR was not highly ranked in structural alignments with BenM owing to significant differences in overall topology and helix lengths. However, the regions of the HTH from TetR and BenM that are involved in DNA interactions act similarly to recognize an ATA DNA duplex. In the TetR–operator complex structure, Gln38 and Pro39 correspond to Gln29 and Pro30 in BenM (Fig. 6*b*; Orth *et al.*, 2000). In members of the TetR family, Pro39 may be replaced by small amino-acid residues (Ramos *et al.*, 2005). Likewise, small amino acids often replace proline at this position in LTTRs (Figs. 2 and 4). In general, it appears to be common for the base-specific recognition of a TA nucleotide duplex to employ a hydrophobic pocket at the terminal end of the interaction helix, especially with proline or valine. Despite the apparent significance of hydrophobic residues in the direct readout of DNA, this phenomenon has not been well characterized.

4. Discussion

The substantial sequence similarity of LTTRs in the DBD region raises questions of how the correct operator–promoter DNA is specifically recognized. For many years, such questions have been posed but have not been answered. For example, nearly twenty years ago it was shown that similar DBDs allow GcvA to substitute for AmpR by activating *ampC* expression, despite the diverse natural functions of these two LTTRs (Everett *et al.*, 1995). Since many bacteria each encode between 40 and 100 LTTRs that have distinct functions, subtle sequence differences must play a role in specificity (Craven *et al.*, 2008). In the full-length proteins the EBDs share relatively low sequence similarity and these regions affect the overall positioning of the DBDs. Furthermore, the EBDs alter the oligomeric conformation when they bind specific effector

molecules. Thus, the relative positions of the DBDs can be dramatically altered by the EBDs and such effects can control promoter recognition.

In BenM, the tetrameric protein shifts from binding sites 1 and 3 to binding sites 1 and 2 of p_{benA} when benzoate and/or muconate are present. However, CatM recognizes similar site arrangements as BenM and the EBDs are nearly identical in structure (Ezezika *et al.*, 2007). At least for these paralogous LTTRs, the EBD-mediated DBD orientation cannot provide the sole answer to the question of how their regulatory effects are distinguished. Instead, additional local promoter features are likely to be important for coupling RNA polymerase binding response to ligand-induced conformational changes of the LTTR. Access of RNA polymerase to the promoter for transcriptional activation may require all of the following: direct readout of bases corresponding to the half-site ATAC, indirect readout of neighboring bases, appropriate binding of the DBD to the DNA (owing to protein–DNA interactions and correct spatial orientation governed by effector binding) and the presence of proper sequences for specific contacts with RNA polymerase. The functional overlap between BenM and CatM provides opportunities to isolate mutants with altered regulation, and the ongoing analysis of selected mutations may help to clarify key features in this system.

The staff of the SER-CAT and SBC-CAT beamlines at the Advanced Photon Source (APS), Argonne National Laboratory provided assistance in data collection. Dr Lirong Chen and Dr John Rose provided local support with the remote data collection. The US Department of Energy, Office of Science, and Office of Basic Energy Sciences under Contract No W-31-109-Eng-38 supported the use of the Advanced Photon Source. AA was supported by the Saudi Arabian Cultural Mission. This research was funded by National Science Foundation grants MCB-0346422 (to CM), MCB-0516914 (to ELN) and IOS-1024108 (to CM and ELN).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981). *Nature (London)*, **290**, 754–758.
Blanchet, C., Pasi, M., Zakrzewska, K. & Lavery, R. (2011). *Nucleic Acids Res.* **39**, W68–W73.
Brennan, R. G. & Matthews, B. W. (1989). *Trends Biochem. Sci.* **14**, 286–290.
Brugarolas, P., Movahedzadeh, F., Wang, Y., Zhang, N., Bartek, I. L., Gao, Y. N., Voskuil, M. I., Franzblau, S. G. & He, C. (2012). *J. Biol. Chem.* **287**, 37703–37712.
Bundy, B. M., Collier, L. S., Hoover, T. R. & Neidle, E. L. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 7693–7698.
Chayen, N. E. (1997). *Structure*, **5**, 1269–1274.
Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G. & Ryu, S. E. (2001). *Cell*, **105**, 103–113.
Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). *Nature (London)*, **364**, 412–420.
Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. & Berman, H. M. (1996). *J. Am. Chem. Soc.* **118**, 509–518.
Clubb, R. T., Omichinski, J. G., Savilahti, H., Mizuuchi, K., Gronenborn, A. M. & Clore, G. M. (1994). *Structure*, **2**, 1041–1048.
Collier, L. S., Gaines, G. L. III & Neidle, E. L. (1998). *J. Bacteriol.* **180**, 2493–2501.
Colyer, T. E. & Kredich, N. M. (1994). *Mol. Microbiol.* **13**, 797–805.
Craven, S. H., Ezezika, O. C., Haddad, S., Hall, R. A., Momany, C. & Neidle, E. L. (2009). *Mol. Microbiol.* **72**, 881–894.
Craven, S. H., Ezezika, O. C., Momany, C. & Neidle, E. L. (2008). *Acinetobacter Molecular Biology*, edited by U. C. Gerischer, pp. 163–202. Norwich: Caister Academic Press.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dock-Bregeon, A. C., Moras, D. & Giegé, R. (1999). *Crystallization of Nucleic Acids and Proteins. A Practical Approach*, 2nd ed., edited by A. Ducruix & R. Giegé, pp. 209–243. Oxford University Press.
Dolan, K. T., Duguid, E. M. & He, C. (2011). *J. Biol. Chem.* **286**, 22178–22185.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Everett, M., Walsh, T., Guay, G. & Bennett, P. (1995). *Microbiology*, **141**, 419–430.
Ezezika, O. C., Collier-Hyams, L. S., Dale, H. A., Burk, A. C. & Neidle, E. L. (2006). *Appl. Environ. Microbiol.* **72**, 1749–1758.
Ezezika, O. C., Haddad, S., Clark, T. J., Neidle, E. L. & Momany, C. (2007). *J. Mol. Biol.* **367**, 616–629.
Gelbin, A., Schneider, B., Clowney, L., Hsieh, S., Olson, W. K. & Berman, H. M. (1996). *J. Am. Chem. Soc.* **118**, 519–529.
Harrison, S. C. & Aggarwal, A. K. (1990). *Annu. Rev. Biochem.* **59**, 933–969.
Harrison, C. J., Bohm, A. A. & Nelson, H. C. (1994). *Science*, **263**, 224–227.
Henikoff, S., Haughn, G. W., Calvo, J. M. & Wallace, J. C. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6602–6606.
Hoffman, M. M., Khrapov, M. A., Cox, J. C., Yao, J., Tong, L. & Ellington, A. D. (2004). *Nucleic Acids Res.* **32**, D174–D181.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Huffman, J. L. & Brennan, R. G. (2002). *Curr. Opin. Struct. Biol.* **12**, 98–106.
Joachimiak, A. & Sigler, P. B. (1991). *Methods Enzymol.* **208**, 82–99.
Jourdan, A. D. & Stauffer, G. V. (1998). *J. Bacteriol.* **180**, 4865–4871.
Kullik, I., Toledano, M. B., Tartaglia, L. A. & Storz, G. (1995). *J. Bacteriol.* **177**, 1275–1284.
Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. (2009). *Nucleic Acids Res.* **37**, 5917–5929.
Monferrer, D., Tralau, T., Kertesz, M. A., Dix, I., Solà, M. & Usón, I. (2010). *Mol. Microbiol.* **75**, 1199–1214.
Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K. & Senda, T. (2003). *J. Mol. Biol.* **328**, 555–566.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (2000). *Nature Struct. Biol.* **7**, 215–219.
Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Nature (London)*, **335**, 321–329.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H., Ebright, R. E. & Berman, H. M. (1996). *J. Mol. Biol.* **260**, 395–408.
Quade, N., Mendonca, C., Herbst, K., Heroven, A. K., Ritter, C., Heinz, D. W. & Dersch, P. (2012). *J. Biol. Chem.* **287**, 35796–35803.
Ramakrishnan, V. (1997). *Annu. Rev. Biophys. Biomol. Struct.* **26**, 83–112.
Ramos, J. L., Martínez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. & Tobes, R. (2005). *Microbiol. Mol. Biol. Rev.* **69**, 326–356.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S. & Honig, B. (2009). *Nature (London)*, **461**, 1248–1253.
Ruangprasert, A., Craven, S. H., Neidle, E. L. & Momany, C. (2010). *J. Mol. Biol.* **404**, 568–586.
Sainsbury, S., Lane, L. A., Ren, J., Gilbert, R. J., Saunders, N. J., Robinson, C. V., Stuart, D. I. & Owens, R. J. (2009). *Nucleic Acids Res.* **37**, 4545–4558.
Schell, M. A. (1993). *Annu. Rev. Microbiol.* **47**, 597–626.
Schell, M. A., Brown, P. H. & Raju, S. (1990). *J. Biol. Chem.* **265**, 3844–3850.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Taylor, J. L., De Silva, R. S., Kovacikova, G., Lin, W., Taylor, R. K., Skorupski, K. & Kull, F. J. (2012). *Mol. Microbiol.* **83**, 457–470.
Tyrrell, R., Verschueren, K. H., Dodson, E. J., Murshudov, G. N., Addy, C. & Wilkinson, A. J. (1997). *Structure*, **5**, 1017–1032.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Watkins, D., Hsiao, C., Woods, K. K., Koudelka, G. B. & Williams, L. D. (2008). *Biochemistry*, **47**, 2325–2338.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, Y., Heath, R. J., Li, Z., Rock, C. O. & White, S. W. (2001). *J. Biol. Chem.* **276**, 17373–17379.
Zheng, G., Lu, X.-J. & Olson, W. K. (2009). *Nucleic Acids Res.* **37**, W240–W246.
Zhou, X., Lou, Z., Fu, S., Yang, A., Shen, H., Li, Z., Feng, Y., Bartlam, M., Wang, H. & Rao, Z. (2010). *J. Mol. Biol.* **396**, 1012–1024.